<u>The Corona Chronologies: Part I. China</u> (was “Analysis_of_Coronavirus-2019_Data_Michael_Levitt.pdf”)

Michael Levitt, Structural Biology, Stanford University School of Medicine, CA, US (13-March-2020)

It is now generally accepted , the COVID-19 epidemic is almost over in China. Today’s analysis uses JHU data (**Table 1**) for the past 49 days to confirm trends we have shown in reports going back to 1-Feb. We separate Hubei from non-Hubei as most cases and deaths have occurred in Hubei (**Fig. 2**). We estimate there will be 3,200 total Hubei deaths and less than 120 non-Hubei deaths in China. There will be 66,000 Hubei cases with a Hubei death rate of 4.5% (1% on Day 0 after being classified as a case; 2.4% on Days 8 & 9, the remaining 1% after day 14, **Fig. 5**). There will be 13,000 non-Hubei cases in China with a death rate of 0.85%. China Non-Hubei deaths seem to occur after 10 days, very similar to the 9 day delay most common for Hubei deaths.

For now, I raise questions for experts who may read this analysis.

(1) Why do most deaths in China tend to occur after 9 or 10 days from infection (Figs. 3,4 & 5)

(2) Why do Hubei cases have a 1% death rate on the day case is confirmed whereas Non-Hubei cases do not (Fig. 5)

(3) Why do China Non-Hubei cases and deaths both peak three days before those in Hubei? Is the explanation in Fig. 6 crazy?

(4) Why do death rates in different countries differ so much (Fig. 7). Do the high rates in Iran mean only the very ill are examined?

(5) Did the epidemic in China slow due to stringent quarantine or rise of immunity in many of those infected but never detected as a case? Can an epidemic be stopped if we use social distancing to contact with fewer people without strict quarantine?

(6) Could certain individuals be naturally immune due to their individual antibody repertoires?

(7) Is what happened on the Diamond Princess a good model for what a world pandemic would be (20% infection rate, 0.04% death rate in over 65-year olds). A big unknown is the role of social distancing on the ship?

The data for China fits a simple sigmoid curve beautifully (see **Fig. 4**) explaining why the early predictions worked out so well. As data accumulates on Non-China cases, we turn our attention to their analysis. This data is very noisy as expected for the early phase and comes from different countries. Preliminary analysis (**Fig. 7**) shows no signs of slowing exponential growth of cases or deaths. More work is needed and we hope that detailed analysis of the epidemic in China will help the rest of the world.

This report will likely be the last on China as we focus our attention on the rest of the world. Actually, the two reports will be concatenated and the Figures and Tables numbered as if one document.

Date	Day	Total Number Cases			Total Number Deaths			Death Rate (%)			Ratio Hubei/ Others	Change Ratio Cases			Change Ratio Deaths			New/Day in Hubei		Non-China	
		Total	Hubei	Others	Total	Hubei	Others	Total	Hubei	Others		Total	Hubei	Others	Total	Hubei	Others	Cases	Deaths	Cases	Deaths
1/22/2020	54	648	565	63	15	15	0	2.28%	2.60%	0.16%	-									6	0
1/23/2020	55	919	672	186	21	20	1	2.29%	2.99%	0.48%	6.2	1.417	1.190	2.961	1.419	1.367	-	107	5	14	0
1/24/2020	56	1216	796	352	28	27	2	2.33%	3.34%	0.48%	6.9	1.324	1.184	1.895	1.348	1.323	1.889	124	7	23	0
1/25/2020	57	1782	1079	591	42	39	3	2.33%	3.61%	0.44%	8.2	1.465	1.355	1.677	1.466	1.462	1.529	283	12	35	0
1/26/2020	58	2556	1480	965	59	55	4	2.32%	3.74%	0.41%	9.0	1.434	1.372	1.632	1.431	1.424	1.538	401	17	48	0
1/27/2020	59	4175	2649	1373	89	83	5	2.12%	3.14%	0.39%	8.0	1.633	1.790	1.423	1.492	1.502	1.350	1169	28	61	0
1/28/2020	60	6353	4309	1807	118	111	7	1.85%	2.58%	0.37%	7.1	1.522	1.627	1.316	1.327	1.334	1.222	1660	28	75	0
1/29/2020	61	8458	5808	2500	143	136	8	1.69%	2.33%	0.31%	7.6	1.331	1.348	1.384	1.219	1.222	1.167	1499	25	90	0
1/30/2020	62	10176	6825	3211	172	163	9	1.69%	2.39%	0.27%	8.8	1.203	1.175	1.284	1.201	1.205	1.130	1016	28	109	0
1/31/2020	63	12263	8150	3990	214	205	9	1.75%	2.51%	0.23%	10.8	1.205	1.194	1.243	1.245	1.254	1.069	1325	42	134	0
2/1/2020	64	15752	10878	4436	275	265	10	1.75%	2.44%	0.23%	10.5	1.284	1.335	1.112	1.285	1.293	1.108	2728	60	157	0
2/2/2020	65	20309	14693	5401	351	339	11	1.73%	2.31%	0.21%	10.9	1.289	1.351	1.217	1.274	1.281	1.107	3814	74	171	1
2/3/2020	66	25282	18893	6127	427	414	12	1.69%	2.19%	0.20%	10.9	1.245	1.286	1.135	1.216	1.221	1.079	4200	75	184	1
2/4/2020	67	30028	22843	7037	494	480	13	1.64%	2.10%	0.19%	11.1	1.188	1.209	1.148	1.157	1.160	1.081	3950	66	200	2
2/5/2020	68	34764	26800	7745	563	549	15	1.62%	2.05%	0.19%	10.8	1.158	1.173	1.101	1.141	1.142	1.105	3957	68	223	2
2/6/2020	69	39233	30528	8509	638	621	17	1.63%	2.04%	0.20%	10.2	1.129	1.139	1.099	1.133	1.132	1.150	3729	73	258	2
2/7/2020	70	43398	34003	8741	720	699	21	1.66%	2.06%	0.24%	8.7	1.106	1.114	1.027	1.127	1.125	1.219	3475	78	299	2
2/8/2020	71	47372	37376	9772	810	783	27	1.71%	2.09%	0.28%	7.6	1.092	1.099	1.118	1.125	1.120	1.306	3373	84	330	2
2/9/2020	72	51054	40531	10254	908	874	34	1.78%	2.16%	0.33%	6.6	1.078	1.084	1.049	1.121	1.117	1.249	3155	92	372	2
2/10/2020	73	54564	43562	10713	1011	971	40	1.85%	2.23%	0.37%	6.0	1.069	1.075	1.045	1.113	1.111	1.179	3031	97	424	2
2/11/2020	74	57814	46373	10825	1086	1041	45	1.88%	2.25%	0.41%	5.4	1.060	1.065	1.010	1.074	1.072	1.129	2811	70	473	2
2/12/2020	75	60937	49114	11216	1189	1137	52	1.95%	2.32%	0.46%	5.0	1.054	1.059	1.036	1.095	1.092	1.157	2741	96	503	2
2/13/2020	76	63950	51785	12123	1342	1283	59	2.10%	2.48%	0.49%	5.1	1.049	1.054	1.081	1.129	1.128	1.147	2671	146	538	3
2/14/2020	77	66664	54189	12451	1520	1455	66	2.28%	2.68%	0.53%	5.1	1.042	1.046	1.027	1.133	1.134	1.108	2404	172	593	3
2/15/2020	78	69044	56275	12696	1655	1585	70	2.40%	2.82%	0.55%	5.1	1.036	1.038	1.020	1.088	1.089	1.065	2086	130	674	4
2/16/2020	79	71178	58146	13018	1768	1694	74	2.48%	2.91%	0.57%	5.1	1.031	1.033	1.025	1.069	1.069	1.061	1871	109	774	5
2/17/2020	80	73213	59956	13247	1880	1800	80	2.57%	3.00%	0.60%	5.0	1.029	1.031	1.018	1.063	1.063	1.071	1811	106	880	5
2/18/2020	81	74742	61297	13445	2000	1914	86	2.68%	3.12%	0.64%	4.9	1.021	1.022	1.015	1.064	1.063	1.080	1341	114	985	6
2/19/2020	82	75655	62049	13606	2125	2031	94	2.81%	3.27%	0.69%	4.7	1.012	1.012	1.012	1.062	1.061	1.092	752	117	1086	8
2/20/2020	83	76216	62387	13829	2212	2111	101	2.90%	3.38%	0.73%	4.6	1.007	1.005	1.016	1.041	1.039	1.079	339	80	1202	11
2/21/2020	84	77147	63006	14140	2309	2202	107	2.99%	3.49%	0.76%	4.6	1.012	1.010	1.023	1.044	1.043	1.059	619	91	1385	14
2/22/2020	85	78187	63677	14510	2402	2288	114	3.07%	3.59%	0.78%	4.6	1.013	1.011	1.026	1.040	1.039	1.060	671	86	1665	19
2/23/2020	86	79027	64142	14885	2512	2389	123	3.18%	3.72%	0.83%	4.5	1.011	1.007	1.026	1.046	1.044	1.082	465	101	2022	26
2/24/2020	87	79637	64372	15266	2606	2472	134	3.27%	3.84%	0.88%	4.4	1.008	1.004	1.026	1.037	1.035	1.089	230	83	2391	36
2/25/2020	88	80452	64758	15694	2703	2558	145	3.36%	3.95%	0.92%	4.3	1.010	1.006	1.028	1.037	1.035	1.080	386	87	2812	45
2/26/2020	89	81503	65189	16314	2765	2608	157	3.39%	4.00%	0.96%	4.1	1.013	1.007	1.039	1.023	1.019	1.087	431	49	3420	56
2/27/2020	90	82756	65570	17186	2818	2645	173	3.41%	4.03%	1.00%	4.0	1.015	1.006	1.053	1.019	1.014	1.098	381	38	4279	70
2/28/2020	91	84270	65944	18326	2875	2683	192	3.41%	4.07%	1.05%	3.9	1.018	1.006	1.066	1.020	1.014	1.112	374	38	5407	86
2/29/2020	92	86145	66379	19766	2937	2724	213	3.41%	4.10%	1.08%	3.8	1.022	1.007	1.079	1.021	1.015	1.110	435	41	6841	105
3/1/2020	93	88248	66800	21449	3006	2763	242	3.41%	4.14%	1.13%	3.7	1.024	1.006	1.085	1.023	1.015	1.137	421	40	8520	134
3/2/2020	94	90477	67080	23397	3081	2800	281	3.41%	4.17%	1.20%	3.5	1.025	1.004	1.091	1.025	1.013	1.159	280	37	10462	172
3/3/2020	95	92767	67217	25550	3166	2836	329	3.41%	4.22%	1.29%	3.3	1.025	1.002	1.092	1.027	1.013	1.172	138	36	12608	220
3/4/2020	96	95258	67337	27921	3254	2870	384	3.42%	4.26%	1.38%	3.1	1.027	1.002	1.093	1.028	1.012	1.167	120	33	14968	274
3/5/2020	97	98208	67464	30745	3353	2901	452	3.41%	4.30%	1.47%	2.9	1.031	1.002	1.101	1.030	1.011	1.175	126	32	17776	341
3/6/2020	98	101823	67577	34246	3456	2931	525	3.39%	4.34%	1.53%	2.8	1.037	1.002	1.114	1.031	1.010	1.163	113	29	21260	414
3/7/2020	99	105803	67657	38146	3600	2959	641	3.40%	4.37%	1.68%	2.6	1.039	1.001	1.114	1.042	1.010	1.220	80	28	25150	530
3/8/2020	100	109736	67706	42030	3785	2985	801	3.45%	4.41%	1.91%	2.3	1.037	1.001	1.102	1.052	1.009	1.249	49	26	29030	690
3/9/2020	101	113923	67738	46186	4013	3006	1007	3.52%	4.44%	2.18%	2.0	1.038	1.000	1.099	1.060	1.007	1.257	32	22	33182	895
3/10/2020	102	119234	67759	51475	4285	3026	1259	3.59%	4.47%	2.45%	1.8	1.047	1.000	1.115	1.068	1.006	1.250	21	19	38465	1147
3/11/2020	103	124261	67771	56490	4529	3042	1487	3.64%	4.49%	2.63%	1.7	1.042	1.000	1.097	1.057	1.005	1.181	12	16	43473	1375
3/12/2020	104	129187	67783	61405	4766	3058	1708	3.69%	4.51%	2.78%	1.6	1.040	1.000	1.087	1.052	1.005	1.149	12	16	48381	1595

Table 1. Showing all we use data for COVID-19. Data, is no longer taken from the WHO website or from jobtube.cn website; it is synced daily to https://github.com/CSSEGISandData/COVID-19/blob/master/csse_covid_19_data/csse_covid_19_time_series/time_series_19-covid-Confirmed.csv the Johns Hopkins University Github repository. All data is smoothed using the LOWESS method (locally weighted scatter-plot smoothing) developed by W. S. Cleveland at Bell Labs in 1985. We divide data into Hubei and non-Hubei as most deaths are in an area centered on Wuhan in Hubei (**Fig. 2**). The death rate is the number of deaths divided by the number of cases confirmed, and Ratio Hubei/Others is the ratio of the death rate for Hubei to the death rate for non-Hubei. The Change Ratio is Value_Today divided by Value_Yesterday. We give the number of new cases and new deaths in Hubei each day (subtracting yesterday from today).

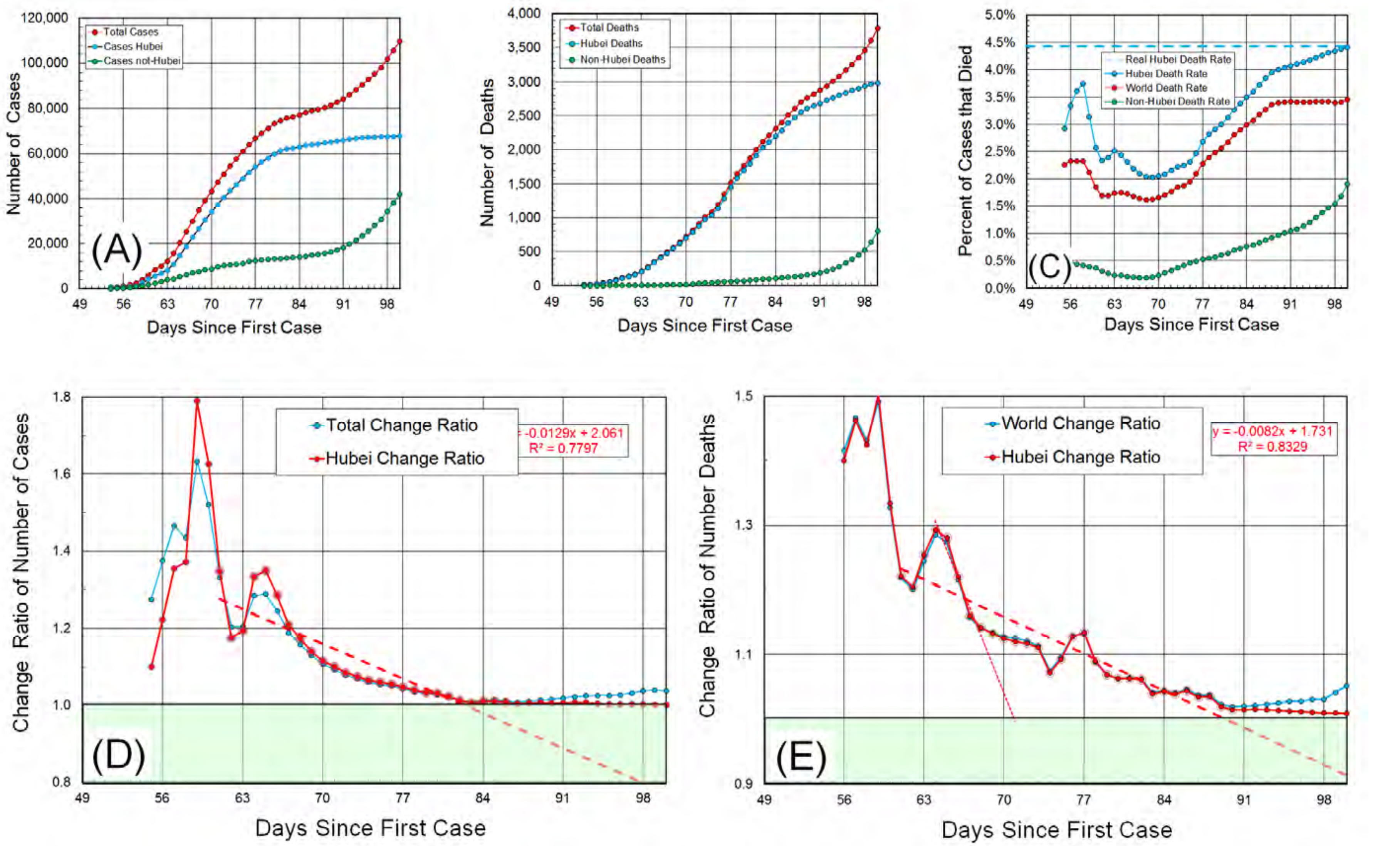


Figure 1. **Variation of COVID-19 data against days since 29 Nov 2019 (guessed date of the first case)**. Data is taken from **Table 1. (A)** shows a slowing increase in number of cases everywhere. **(B)** confirms that almost all the deaths are in Hubei. **(C)** shows that the Hubei death rate initially decreased from 2.5% on 27-Jan. to 1.9% on 7-Feb. only to rise to 4.0% today. Such a rise of the Hubei death rate in (C) makes no sense as the virus is not becoming more virulent. This discrepancy arises because all deaths do not occur on the same day a case is diagnosed. A proper death rate distribution gives a real Hubei death rate of 4.7% (**Fig. 5**). **(D)** and **(E)** show that the change ratio in total cases or deaths (Value_Today / Value_Yesterday) is decreasing steadily. In (D) & (E) we add linear trend-lines using data from 1/29/2020. The change ratio for cases and deaths is an excellent fit to a straight line. In (E) we also show a red short-dashed line of the linear fit to the four data points for 29-Dec to 01-Feb; this trend was seen in the first draft of this analysis dated 2/2/20, giving rise to the hope I expressed that the growth of deaths would slow soon.

			16-Feb				12-Feb				6-Feb				4-Feb				2-Feb				31-Jan		
Province or City in Hubei	Population	Deaths / million pop	Cases	Deaths	Death Rate	Death Ratio	Cases	Deaths	Death Rate	Death Ratio	Cases	Deaths	Death Rate	Death Ratio	Cases	Deaths	Death Rate	Death Ratio	Cases	Deaths	Death Rate	Death Ratio	Cases	Deaths	Death Rate
Hubei	58,500,000	29.0	58,182	1696	2.91%	1.44	34,874	1176	3.37%	1.90	22,112	618	2.79%	1.29	16,678	479	2.87%	1.37	11,177	350	3.13%	1.41	7,153	249	3.48%
Wuhan	11,080,000	118.1	41,152	1309	3.18%	1.45	19,558	902	4.61%	1.89	11,618	478	4.11%	1.32	8,351	362	4.33%	1.37	5,142	265	5.15%	1.38	3,215	192	5.97%
Huanggang	7,403,000	10.5	2,831	78	2.76%	1.34	2,441	58	2.38%	1.81	1,897	32	1.69%	1.28	1,645	25	1.52%	1.47	1,246	17	1.36%	1.21	726	14	1.93%
Xiaogan	4,900,000	14.3	3,279	70	2.13%	1.43	2,839	49	1.73%	1.96	2,141	25	1.17%	1.39	1,462	18	1.23%	1.29	918	14	1.53%	1.17	628	12	1.91%
Jingzhou	3,692,000	10.0	1,501	37	2.47%	1.61	1,114	23	2.06%	2.30	885	10	1.13%	1.11	713	9	1.26%	1.50	499	6	1.20%	1.50	287	4	1.39%
Ezhou	1,050,000	33.3	1,274	35	2.75%	1.17	1,010	30	2.97%	1.67	471	18	3.82%	1.00	382	18	4.71%	1.20	306	15	4.90%	1.67	227	9	3.96%
Jingmen	3,023,000	10.9	915	33	3.61%	1.38	725	24	3.31%	1.41	553	17	3.07%	1.06	422	16	3.79%	1.45	345	11	3.19%	2.20	251	5	1.99%
Suizhou	2,500,000	9.6	1,267	24	1.89%	1.71	1,160	14	1.21%	1.56	915	9	0.98%	1.13	706	8	1.13%	1.60	458	5	1.09%	5.00	304	1	0.33%
Yichang	4,060,000	5.9	895	24	2.68%	2.18	810	11	1.36%	1.57	610	7	1.15%	1.75	496	4	0.81%	4.00	392	1	0.26%	1.00	276	1	0.36%
Xiangyang	900,000	22.2	1,155	20	1.73%	1.54	1,101	13	1.18%	4.33	838	3	0.36%		735	2	0.27%		548	0	0.00%		347	0	0.00%
Xiantao	1,175,000	16.2	531	19	3.58%	1.19	478	16	3.35%	3.20	307	5	1.63%	1.25	225	4	1.78%	1.33	169	3	1.78%	3.00	97	1	1.03%
Huangshi	2,450,000	6.1	983	15	1.53%	1.67	899	9	1.00%	4.50	635	2	0.31%	1.00	509	2	0.39%	1.00	334	2	0.60%	1.00	209	2	0.96%
Tianmen	1,731,000	5.8	485	10	2.06%	1.00	336	10	2.98%	1.00	163	10	6.13%	1.00	128	10	7.81%	1.00	115	10	8.70%	1.43	82	7	8.54%
Xianning	2,800,000	3.6	861	10	1.16%	1.43	528	7	1.33%		443	0	0.00%		384	0	0.00%		296	0	0.00%		206	0	0.00%
Qianjiang	1,000,000	6.0	182	6	3.30%	1.20	94	5	5.32%	5.00	74	1	1.35%	1.00	54	1	1.85%	1.00	35	1	2.86%	1.00	27	1	3.70%
Enshi	750,000	5.3	249	4	1.61%	1.33	210	3	1.43%		157	0	0.00%		138	0	0.00%		111	0	0.00%		87	0	0.00%
Shiyan	3,340,000	0.6	612	2	0.33%	2.00	559	1	0.18%		395	0	0.00%		318	0	0.00%		256	0	0.00%		177	0	0.00%
Shennongjia	76,000	0.0	10	0	0.00%		10	0	0.00%		10	0	0.00%		10	0	0.00%		7	0	0.00%		7	0	0.00%

Table 2. Number of cases, number of deaths, death rates and change ratios in death numbers (death ratio) shown for 17 Hubei cities from 31 Jan to 16 Feb. City data is sorted by decreasing number of deaths. We distinguish death rates ≥ 3% (scarlet), ≥ 1% (rose) & < 1 % (green). The deaths per million population is much higher in Wuhan than any other city at almost 120 per million (0.012%). The number of cases (clinically plus laboratory diagnosed) is 0.37% of the Wuhan population of 11 million. On 31-Jan. there were 8 of 17 cities with death rates less than 1%; by 16-Feb., there were only 2 of 17.

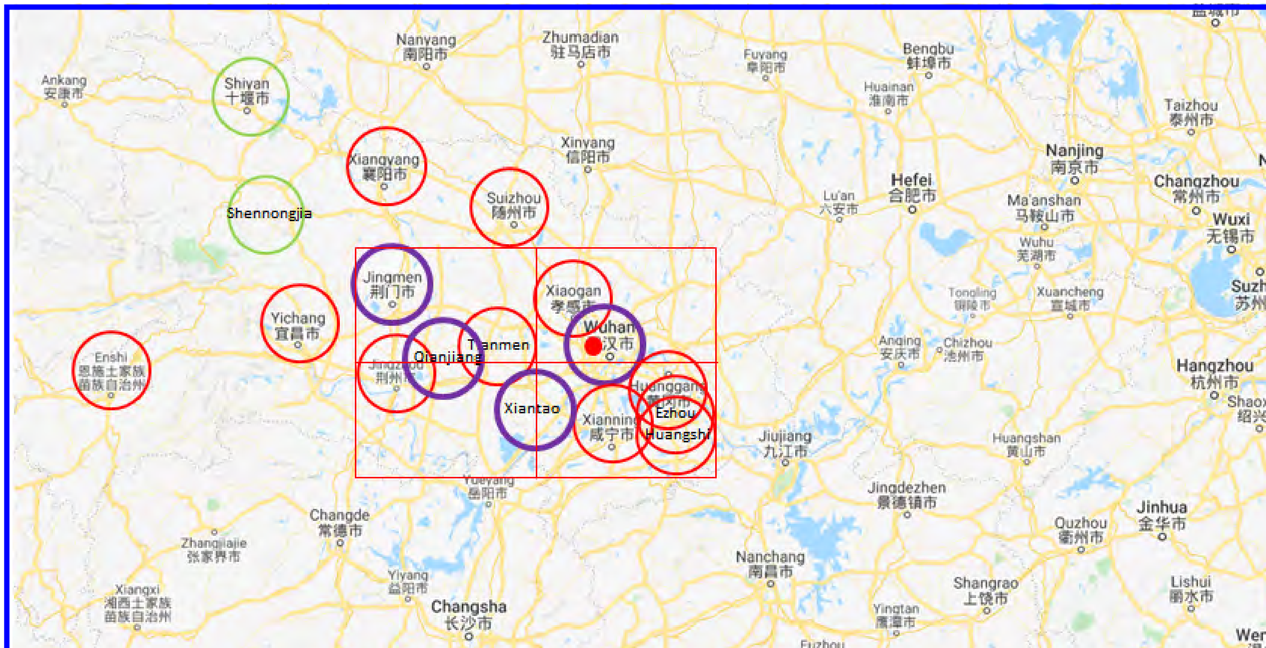


Figure 2. Map of Hubei circling in purple cities with a death rate of ≥ 3%, in red cities with a death rate of ≥ 1% and in green other cities for which there is data in Table 2. Most deaths are localized to a 90 km x 35 km area centered near Tianmen and high death rates occur in four cities: Wuhan, Jingmen, Qianjiang and Xiantao (See **Table 2**). Two cities, in the same area have low death rates, comparable to those elsewhere in China and the rest of the world (data from jobtube.cn from 31-Jan. to 16-Feb.). The red dot marks the Wuhan South China Seafood Market thought to be the source of this coronavirus.

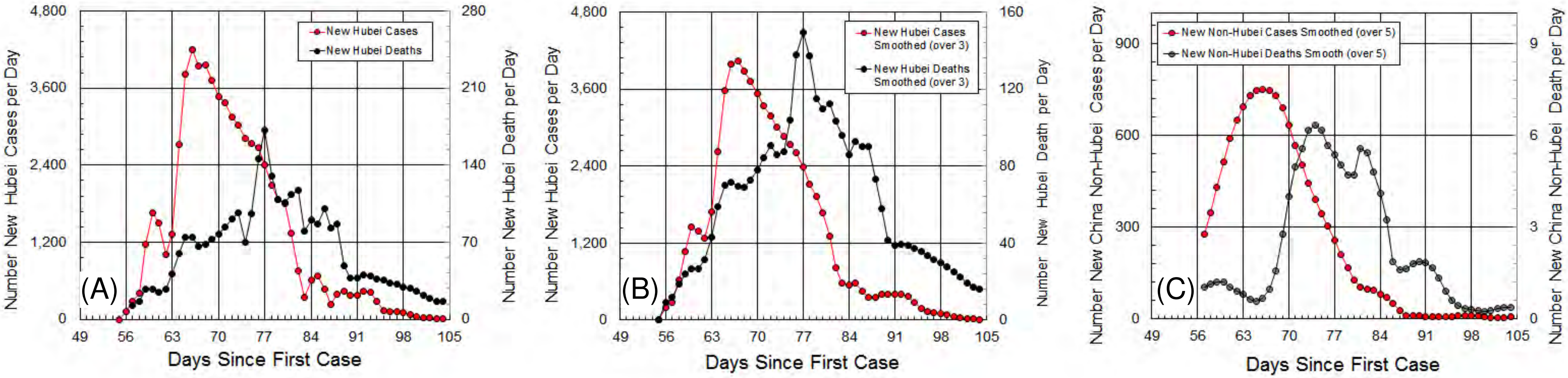


Figure 3. Time variation of number of new cases and new deaths in China, separated into Hubei and elsewhere in mainland China (Non-Hubei).

(A). Showing the number of new Hubei cases per day (red line) and the number of new Hubei deaths per day (black line).

(B) The same data is smoothed by averaging over a three-day window so that, for example, the value plotted on day 69 is the average of the values on days 68, 69 & 70. These smoothed curves clearly show that the number of new Hubei cases per day peaked on Day 69 (6-Feb) and that the number of new Hubei deaths per day peaked on Day 78 (15-Feb.), which is 9 days later. For sigmoid growth like that shown in **Fig. 4**, the number of new cases or deaths reaches a maximum midway through the curve. This predicts the total number of Hubei cases will reach 60,000 (laboratory plus clinically diagnosed cases), approximately twice 28,208, the number of such cases on 6-Feb. This also predicts total number of Hubei deaths will reach 2,914, twice 1,457, the number of Hubei deaths on 15-Feb. Better analysis in **Fig. 4** gives asymptotic values of 65,834 and 3,150 for number of Hubei cases and deaths, respectively.

(C) Showing the variation with time of the smoothed number of new Non-Hubei cases in China per day (red line). Although smoothed by averaging over a window of five values, this data remains noisy. Nevertheless, it does indicate that a peak in the number of new Non-Hubei cases in China occurred on day 67 or 68 (4-Feb. or 5-Feb.) allowing the maximum total number of Non-Hubei cases to be estimated as twice 7,037 or 7,745, the values on 4-Feb. or 5-Feb, for a value of between 14,000 and 16,000. The same argument estimates the total number of non-Hubei deaths to reach a 160. Again, **Fig. 4** gives better asymptotic values of 13,075 and 109 for the total number of Non-Hubei cases and deaths, respectively.

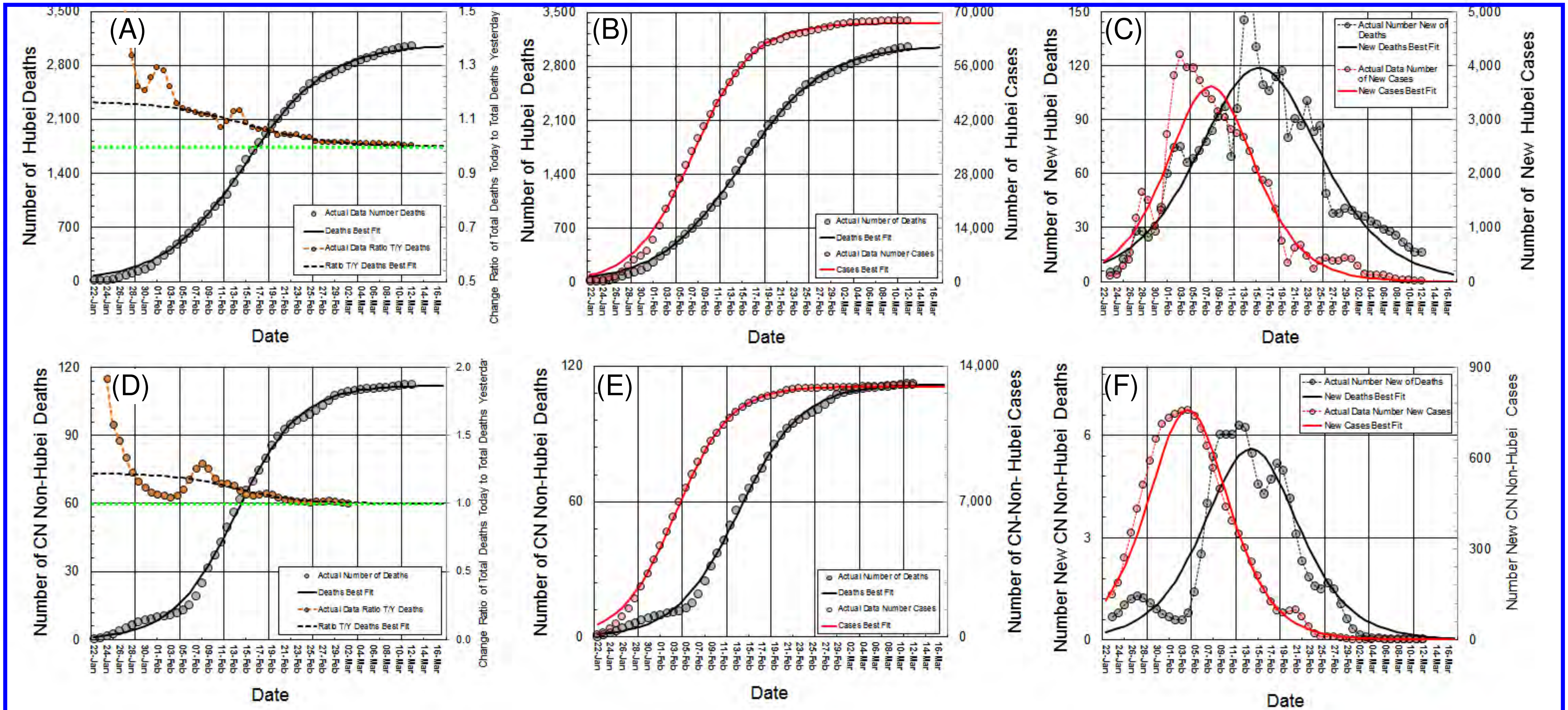


Figure 4. Fit of a sigmoid function to the total number of COVID-19 cases and deaths in Hubei. (A) The best fit (black line) to the actual deaths (black dots). The fit is obtained using Excel Solver to find parameters *A, B & C in f(x) = A/(1+exp(-(x-B)/C))* that minimize the weighted RMS difference of calculated and actual number (weight=sqrt(number deaths). We calculate ratio of value today to those yesterday (T/Y, black dashed line) and compare with the actual data (orange dashed line and circles on secondary axis). The fit is excellent and the calculated ratio decreases approximately linearly towards a value of 1.0 as assumed in **Fig. 2 (E)**. =A/(1+EXP(-(x-B)/C)) **(B)** Sigmoid fits to both the number of cases and number of deaths in Hubei. The final total number of Hubei cases will be close to 66,000, while the current estimate for total number of deaths will be close to 3,200. This will mean an overall Hubei death rate of almost 5% (3,053/67,067=4.55%). **(C)** By subtracting values for yesterday from today, the sigmoid function fitted to the actual number of new Hubei cases or deaths shown in **Fig. 4B**, gives the number of new Hubei cases or new Hubei deaths (solid red and black lines, respectively). These curves are a good fit to the actual number of new Hubei cases or deaths (red and black transparent circles joined by dashed read and black lines, respectively), although the real data is noisy with large fluctuations. The smooth new cases curve (solid red line) peaks at Day 70.4 and the smooth new deaths curve (solid black line peaks at Day 78.6). Corresponding plots for cases and deaths in China but Non-Hubei is plotted in panels **(D)**, **(E)** & **(F)**. The Non-Hubei death rate is almost 1% (122/12939=0.86%), which is about 5 times lower than that in Hubei. The for sigmoid curve parameters (*A, B, C)* = (67165, 70.3, 4.65) for Hubei cases, (3071, 77.9, 6.47) for Hubei deaths, (12956, 66.5, 4.26) for Non-Hubei cases, and (112, 76.4, 5.00) for Non-Hubei deaths.

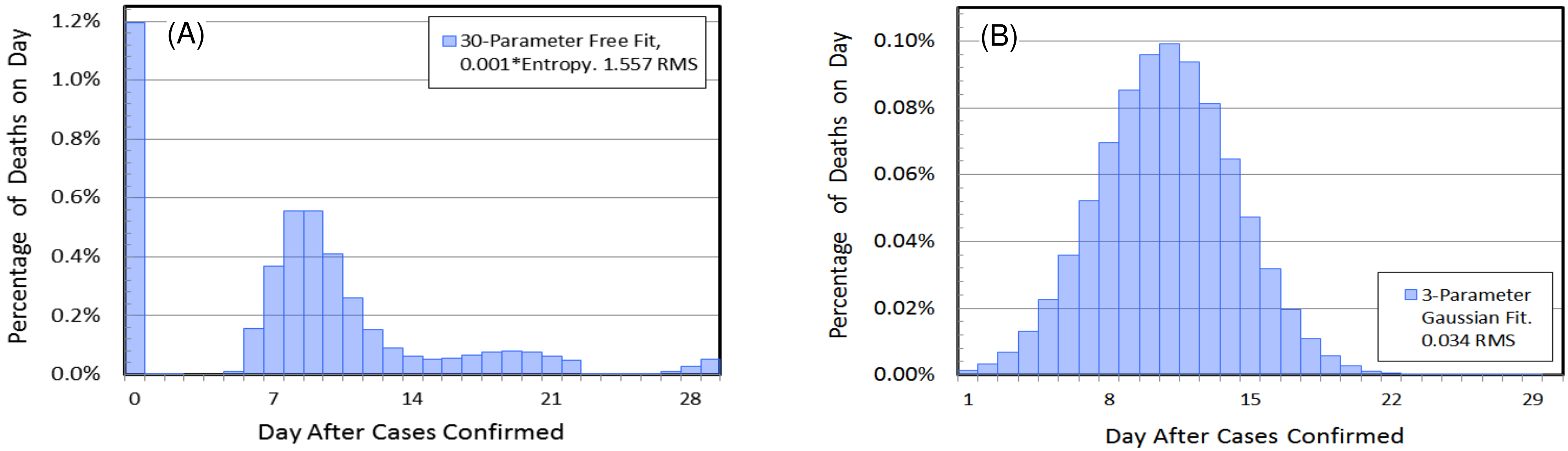


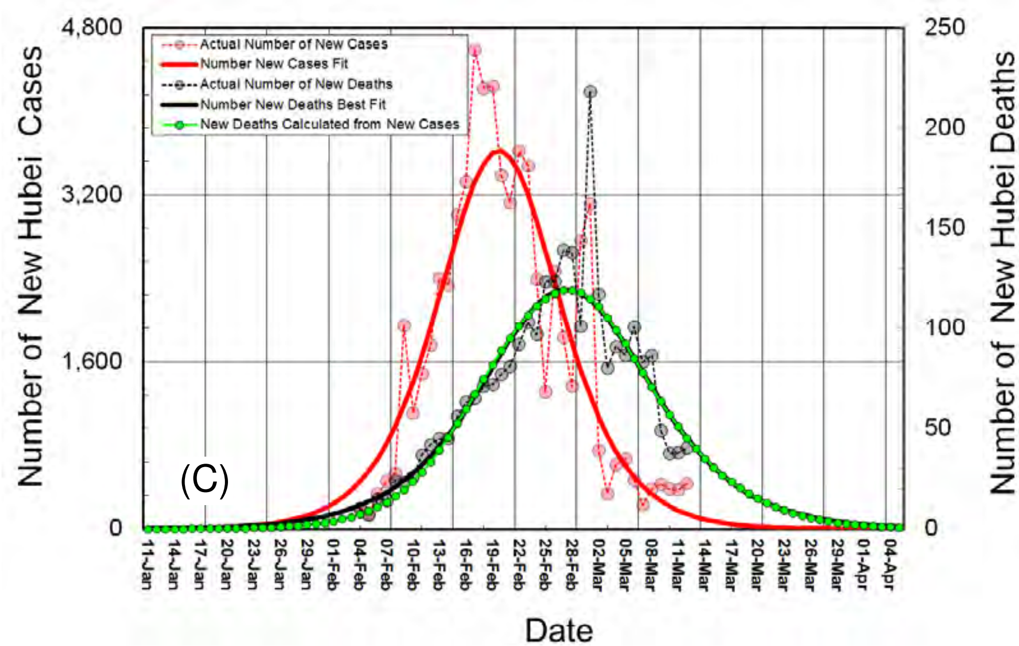
Figure 5. Relating new cases to new deaths via a death rate distribution, which gives the fraction of cases that die *i* days after a case is confirmed. If P_i is fraction of cases that die after i = days, the number of new deaths on day n, D_n, is the sum of deaths from the new cases, C_{n-i}, on previous days, where $D_n = C_n{*}P_0 + C_{n-1}{*}P_1 + C_{n-2}{*}P_2 + \ldots + C_{n-29}{*}P_2$. The total death rate is ΣP_i. Excel Solver is used to determine values for P_n unknowns in two ways:

(1) 30 parameters, one for each P_n value.

(2) 3 parameter Gaussian $P_n = P{*}exp(-((n-Q)/R)^2)$.

The 30 parameters distributions are smoothed with an entropy penalty of $-W\Sigma P_n ln(P_n n)$ added to the weighted least squares fit of predicted and actual number of new deaths.

(A) The death rate distribution that best fits predicted deaths to actual deaths in Hubei has 4 peaks with a death rate of 1.2% on day 0, the day a case is confirmed, of 2.7% summed over days 5 & 15 (centered on Day 8.5) and about 0.6% over later days. The entropy weighty = 0.001. the total death rate in all cases is 4.43%.



(B) A death rate distributions allows China Non-Hubei new deaths to be predicted from China Non-Hubei new cases. Single Gaussian fit gives a broad peak centered on Day 10. The total death rate 0.84%. For both Hubei and Non-Hubei, the death rate is higher than in **Fig. 1C**.

(C) The new deaths predicted from actual new cases (red line) is shown as a green dotted line. The fit between the predicted new deaths and the actual new deaths (black line) is excellent (hiding black line) except for 15-Jan. to 29-Jan. when it is low. In that period, the number of new cases confirmed could have been underestimated due to difficult conditions in Hubei.

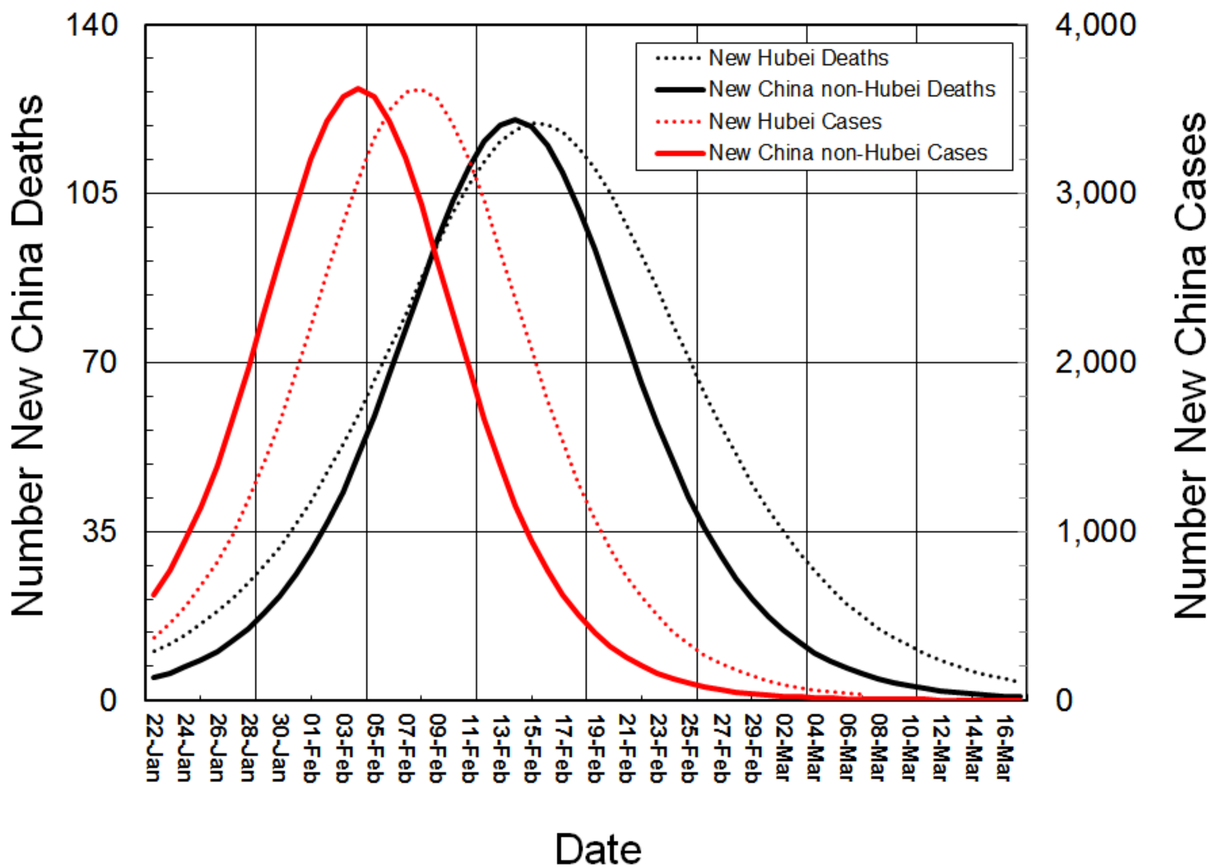


Figure 6 compares the sigmoid curves for cases and deaths in Hubei and China Non-Hubei (see **Fig. 4**) . The smaller number of Non-Hubei cases are scaled by a factor of 4.99 so they are the same height as Hubei cases. The same is done to the much smaller number of Non-Hubei deaths, which are scaled by a factor of 20.06. This shows that Non-Hubei cases peaked three days before those in Hubei, while Non-Hubei deaths peaked two days before those in Hubei. This seems impossible but I believe it may be explained if the Non-Hubei cases were all infected in Hubei three days before the majority of those infected in Hubei. This means that these Non-Hubei cases are from infected people who left Wuhan for the Spring Festival (Chinese New Year) and before the city was locked down on 23 Jan. The lack of further infection suggests that the quarantine of those coming from Hubei to other parts of China prevented any further spread of infection. This conjecture is still uncertain but illustrates just how much analysis of the data may reveal.

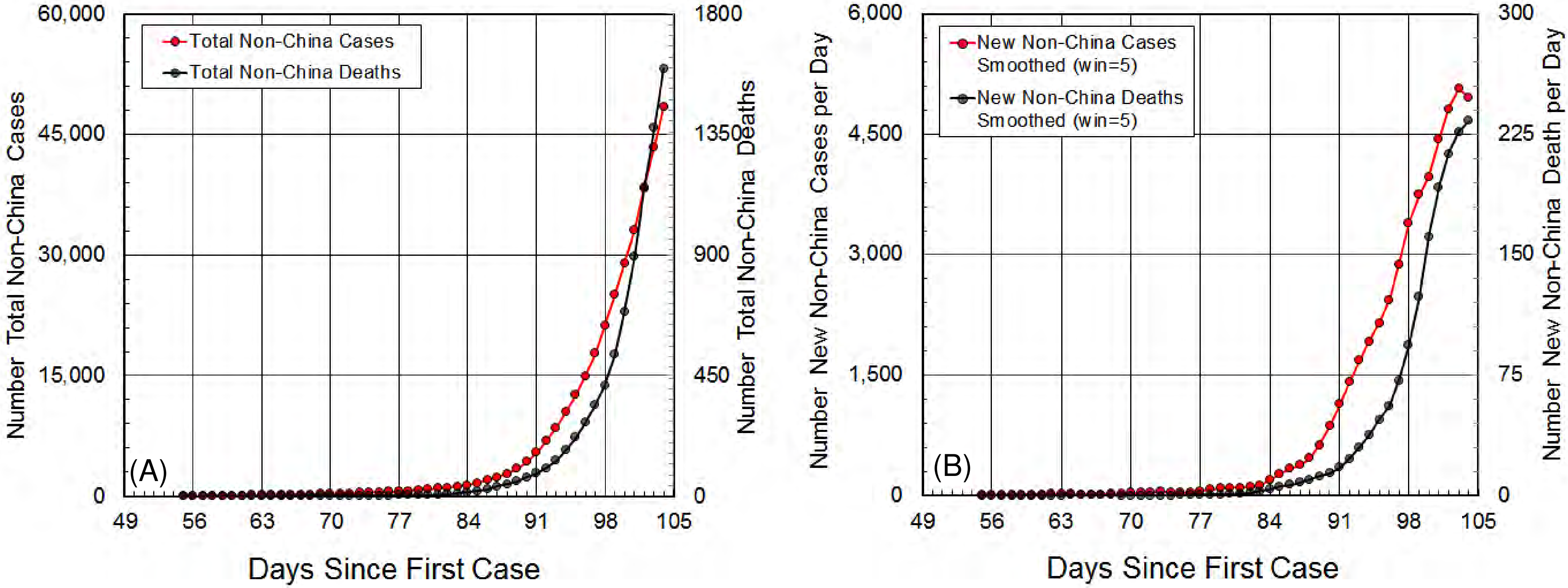


Figure 7 shows the number of cases and deaths outside China. These plots involve small numbers and are beset by high levels of noise. Still, now is the time when prediction is important.

(A) Shows both cases and deaths are increasing rapidly.

(B) Shows that the number of new cases and new deaths per day are increasing together and without the lag seen in **Fig. 4B & E**. This suggests that many cases are not detected until symptoms are severe and patience die on the same day.

Data on cases outside of China is now analyzed in Part II of these reports that follows below.

The Corona Chronologies: Part II. The Rest of the World

Michael Levitt, Structural Biology, Stanford University School of Medicine (14 March 2020)

Since 1 Feb-2020, I have been analyzing the coronavirus COVID-19 epidemic. The last report on China was entitled "30.Analysis_of_Coronavirus-2019_Data_Michael_Levitt.pdf". It was distributed on 1-March by links in Dropbox: (http://bit.ly/2OsE5Sf) and Github (https://csblab.github.io/novoCoronavirus-Analysis/), as WhatsApp, WeChat and email to selected friends and colleagues. Doing all this has been a strain on limited human resources so I am now changing direction and focusing on 'Non-China' or the Rest of the World. I now have a more experience, but early prediction the path and outcome of the world epidemic in not going to be easy" the numbers are always small at the beginning, the definition of what constitutes a case is vague and likely differs from country to country.

This first report consists of four sections: (1) A lay-persons introduction taken from a radio show appearance on 2 March. (2) A mathematical exercise (likely high-school level, but proud to have solved it), which gives a straight-line relationship between growth rate and number of occurrences (e.g. the percent growth in cases per day should depend linearly on the total number of cases to date with a slope related to the sharpness of the sigmoid curve). (3) Application of this relationship to the epidemics in Hubei and China Non-Hubei, which are now essentially over. (4) Application of this relationship to the epidemics in the rest of the world, both in aggregate and to countries with enough cases and deaths. This shows the very first signs of improvement: there is a decreasing case growth rate in South Korea and Italy and a peak in new cases in South Korea.

A Simple Explanation of Viral Epidemics (Brian Kilmeade, Fox Radio 3-Mar-2020)

I am not trained or experienced in the area either of virology or epidemiology. Still, I have been thinking about coronavirus non-stop for the past 30 days. I think I start to have a pretty good grasp about what is going on. Let me start with a simple analogy. If each person infects about 2 others (actual value, known as R_0, is closer to 2.2), then if they are infective for 3 days, that same person will infect on average 1.3 people a day. This means that the number of people infected will grow by 30% a day.

This is exponential growth and it is very fast. If it was the interest you got from your bank or a great investment, then $1 would become $2,620 in 30 days and 17 billion dollars after 90 days. If there was as single person infected by Coronavirus, then at 30% growth a day, it would take 75 days to infect all the US population and just another 12 days to infect the entire world.

This is fast and wonderful if it is money but really scary if it is people infected. In fact no bank will pay 30% interest a day and in the real world viral infections do not grow exponentially for long. Something slows the growth down and signs of this something can be detected quite early on. We will discuss these slowing factors in a bit.

For now something about my personal involvement. By 30 January there were already 10,000 cases and 170 deaths in China and the number of cases and deaths was growing at 30% a day. It seemed like a doom's-day scenario. Looking closely, showed that the rate of growth was not fixed as it would be for exponential growth, instead it was decreasing from 29% to 25% to 22% for numbers of deaths on 30th , 31st January and 1st February.

I knew very little about epidemiology at that time but these decreasing numbers seemed to give hope as it is obvious that when the daily growth in deaths drops to 0%, the infection is over. I was very excited to find this and contacted friends to tell them that the end of the world was not close, even in China. They were really happy and someone (I know not whom) translated my two page analysis into Chinese and posted it onto Chinese social media.

The reaction was overwhelming and I now had to keep on doing the analysis to see if what I had predicted was actually happening. The numbers behaved well: new cases in China peaked on 7th February and new deaths peaked on 16th February, nine days later. This allowed me to quiet accurately predict the eventual number of cases in China as 80,000 and eventual number of deaths as 3,500. I also noticed very early on that the death rate was ten times higher much Hubei than in the rest of China. This difference has dropped so that now I know that Hubei has a 5% death rate compared to 1% for the rest of China.

It is still unclear why the increase in cases and deaths got slower. It could be due to immunity of others who were sick and recovered or who were infected but never showed symptoms while still developing antibodies. It could be social distancing. It could be washing hands well and using a mask if sick. It could be all these things together. Key is that these factors reduced the number of people a sick person infects from 2.2 to below 1.0, which will stop the exponential growth and the epidemic.

Over the past few days, I have tried to see if the prediction methods I used in China a month ago, can be used to say something about the growth of Coronavirus in the rest of the world.

Brian Kilmeade Show 3-2-2020 from minute 20:20 to minute 36:00.
https://podcasts.google.com/?feed=aHR0cDovL2ZlZWRzLmZveG5ld3NyYWRpby5jb20vYm5q&episode=YmEyNTA1YmYtMzIwYy00YmY3LWE4YWItYWI3MjAxMTIjNTVl&hl=en-IL&ved=2ahUKEwjWjOyhhP_nAhUP3qQKHY_DCWoQjrkEegQILxAE&ep=6

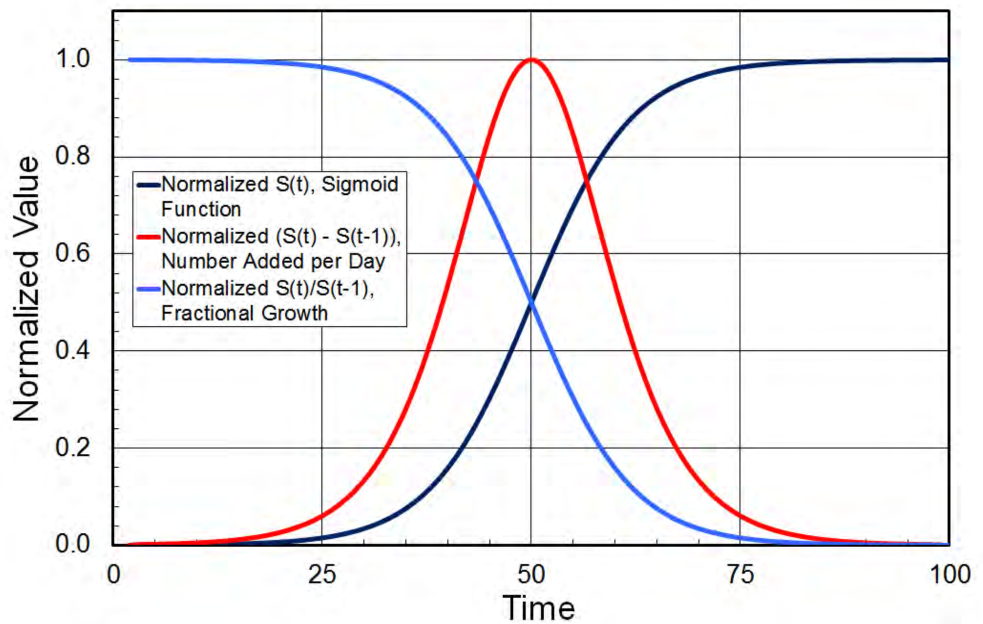


Figure 8. The change with time *t* of function S(t) = A/(1+exp(-(t-B)/C)), of the total number of cases or deaths (black line); the daily growth, *S(t)-S(t-1)* in red; and the growth ratio *S(t)/S(t-1)* in blue. All the values are normalized to be between 0 and 1. This normalization, makes the *A* parameter irrelevant. The parameter *B* is set to 50 days (the midpoint) and the parameter C is set to 6 days (the width).

A Straight-line Relationship Between Growth and Number

In my first report I assumed that the rate of growth (or change ratio in number after one day) depended linearly on time. This assumption ended up being wrong as seen from **Fig. 8** where the growth rate (blue line) is flat at first, then drops more-or-less linearly and finally flattens out again. This dependence is not linear over all the range.

I just found that for sigmoid function *S(t)*, the growth measured by *S(t)/S(t-1)* (time is *t* today and *t-1* yesterday) depends on *S(t)*, the value of the sigmoid function itself (**Fig. 9**). This means that the change ratio (or growth) in number of cases should decrease linearly as the number of cases increases and the change ratio in deaths should decrease linearly as the number of deaths increases (**Fig. 10**).

Consider the Sigmoid function S_t that goes from $t = 0$ to 2B.

$$S_t = A / (1 + e^{-(t-B)/C})$$

$$A / S_t - 1 = e^{-(t-B)/C}$$

$$(A - S_t) / S_t = e^{-(t-B)/C} \quad \text{.................. Eqn. (1)}$$

and

$$(A - S_{t-1}) / S_{t-1} = e^{-(t-1-B)/C} \quad \text{.................. Eqn. (2)}$$

Divide Eqn. (2) by Eqn. (1) and simplify

$$(A - S_{t-1}) / S_{t-1} = e^{1/c} e^{-(t-B)/C}$$

$$\frac{(A - S_{t-1}) S_t}{(A - S_t) S_{t-1}} = e^{1/C}$$

$$(A - S_{t-1}) S_t = e^{1/C} (A - S_t) S_{t-1}$$

$$AS_t - S_{t-1} S_t = e^{1/C} (AS_{t-1} - S_t S_{t-1})$$

Now divide both sides by S_{t-1}

$$\frac{AS_t}{S_{t-1}} - S_t = e^{1/C} (A - S_t)$$

$$\frac{AS_t}{S_{t-1}} = e^{1/C} (A - S_t) + S_t$$

$$\frac{AS_x}{S_{x-1}} = Ae^{1/C} - e^{1/C} S_x + S_x$$

$$\frac{S_x}{S_{x-1}} = e^{1/C} - (e^{1/C} - 1) \frac{S_x}{A}$$

Figure 9. Proving the linear relationship. The algebra above shows that the growth ratio or change ratio depends linearly on the value of the sigmoid function.

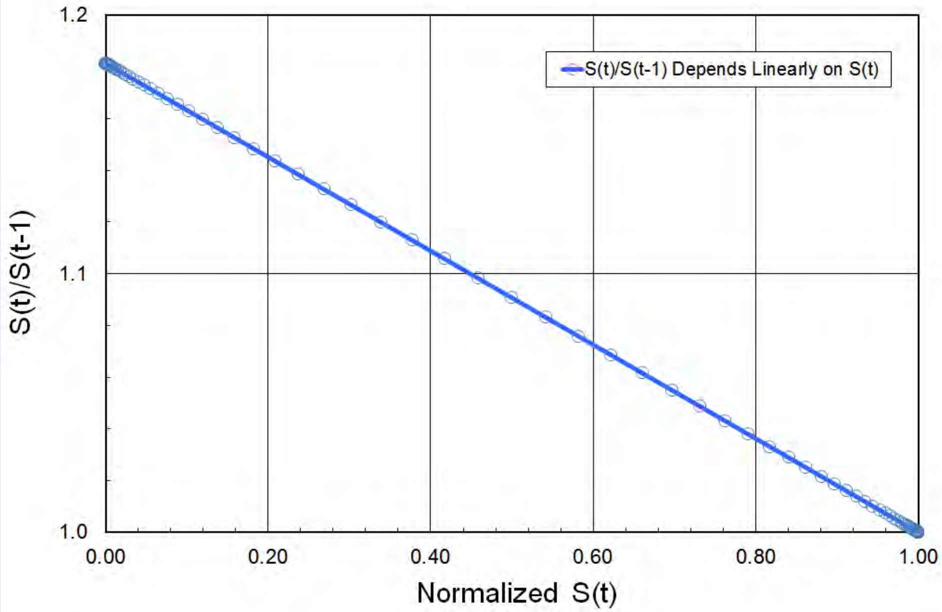


Figure 10. Showing that the growth ratio *S(t)/S(t-1)* (or fractional change) depends linearly on the *S(t)* value at time *t*, The straight line has a slope of *–exp(1/C)*, where parameter *C* is a measure of the width of the transition (See **Fig. 1**).

<u>Revisiting Deaths and Cases in China, Hubei and Non-Hubei</u>

We revisit the plots of the Change Ratio shown in **Fig. 1** (D) & (E). The number of cases and deaths are smoothed by the LOWESS method (**Table 1**) values before calculating Change Ratios and plotted **against** total number of deaths or cases rather than against time. The results in **Fig. 11** show a clear tendency for the Change Ratio to drop. The smooth drop of Change Ratio for cases in Hubei towards the straight line is suspect and needs further investigation (**Fig. 11B**).

Fig. 12 shows that while South Korea is well past the mid-point, both Italy, Iran may be close to their midpoints.

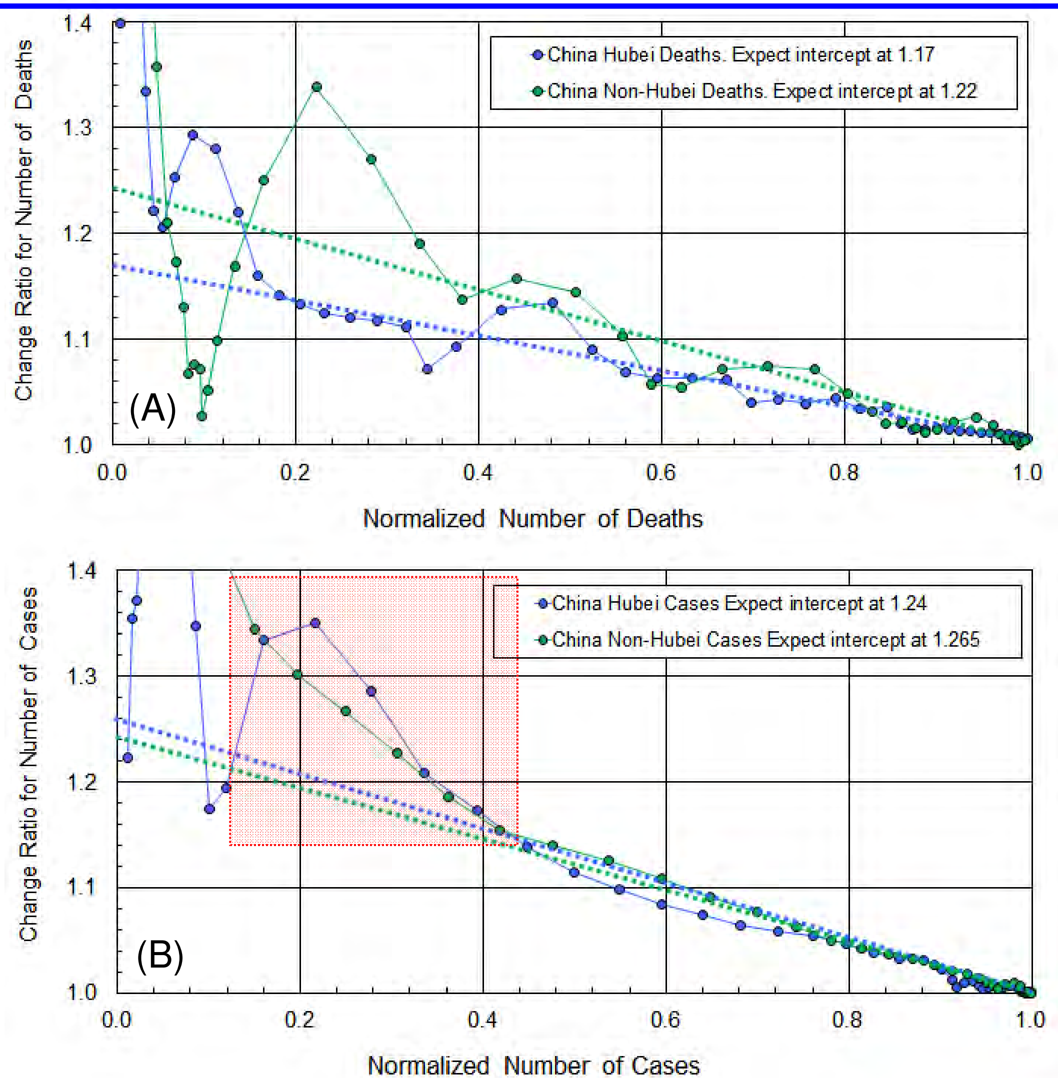


Figure 11. Showing how the Change Ratio (Number_Today divided by Number_Yesterday) fits the straight lines predicted by the equations in Fig. 9 for both number of deaths (Panel (A)) and number of cases (Panel (B)). The lines drawn are not best linear fits but rather the dependence predicted by the equation in **Fig. 9**, using the C parameter values from the best sigmoid fit to the data.

Unexpectedly, the curve for China, non-Hubei cases is particularly smooth and indicates a non-linear dependence that needs to be understood. Either the algebra in **Fig. 9** is wrong or the sigmoid function does not account for the rapid drop is Change Ratio seen for Hubei. This is very important as the rapid drop of the initially high change ratio is what allowed these epidemics to be controlled.

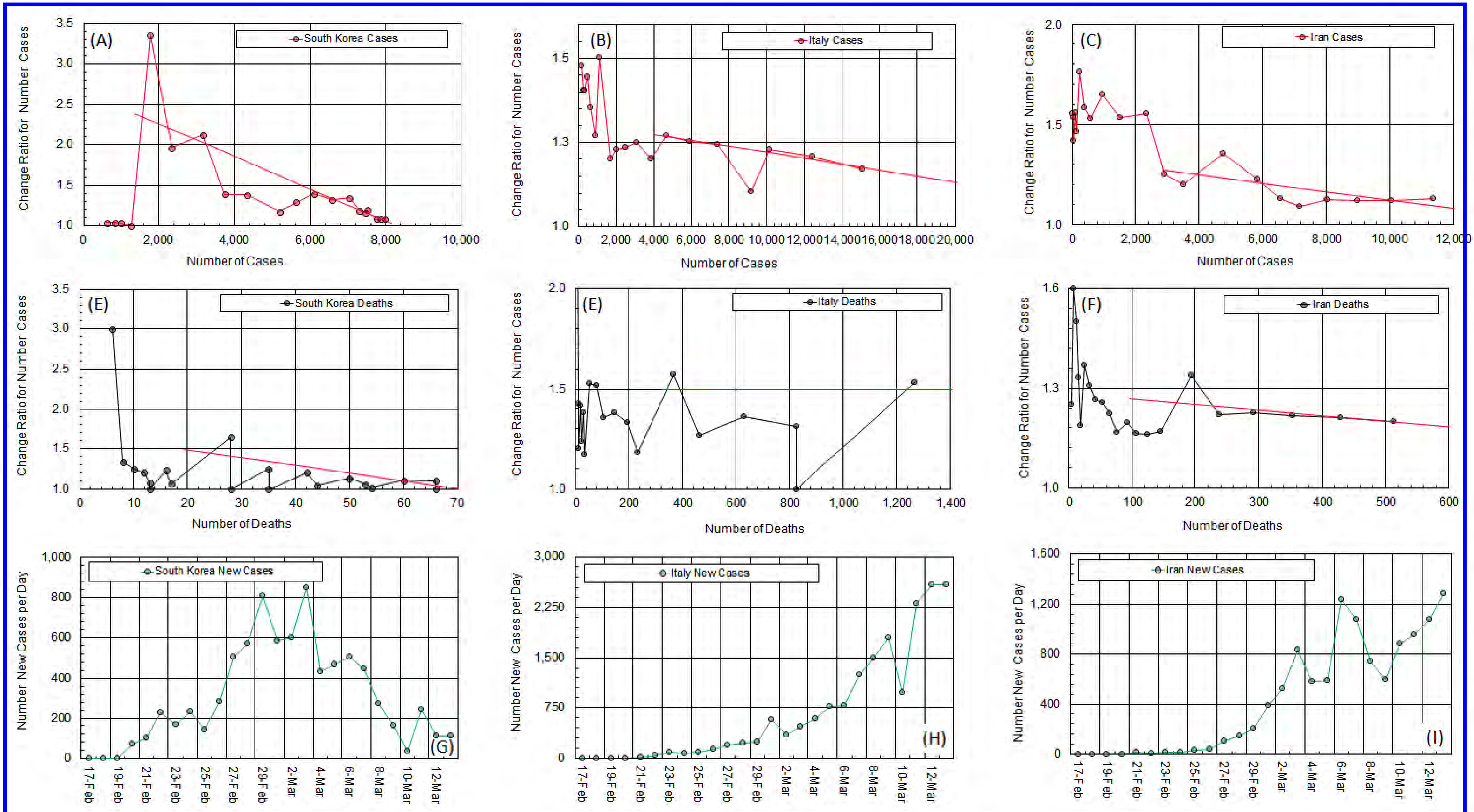


Figure 12. Showing Change Ratios of cases vs. number of cases in panels (A) to (C) for South Korea, Italy and Iran, respectively. There is a clear decreasing trend of the Case Change Ratio for South Korea, and Iran and Italy. Panels (E) to (H) show that same plot for the Death Change Ratio. All trends are weaker except for South Korea. Plots of number of new cases against date, show that South Korea has peaked. There is a double peak for Iran while that is Italy may have peaked but we need a few more days of data to be sure. Notice the false peak that excited me on 9-Mar.

World Population Death Rate

The case fatality rate is the total number of deaths divided by the total number of cases at the end of the epidemic. Its value depends very much on how a caser is defined. Thus **Figs. 1** and **XXX** that the case fatality rate in Hubei is 5% whereas in China outside Hubei is it less than 1%. Assuming that we are dealing with the same virus, this difference is likely due to how case is defined. Given the very difficult conditions in Wuhan, it is understandable and expects that only the most severe cases would be counted. This hypothesis gains weight when we see that comparison of the case sigmoid curve with the death sigmoid curve for Hubei shows that about ¼ of the fatalities occurred on the same day a case was confirmed. This did not happen in China outside Hubei, when most of those who died did do between 8 and 10 days after being confirmed as a case (Fig. XXX).

Given this difficulty we try here to estimate the population fatality rate for the one epidemic that had both many cases and a high percentage of cases: the Diamond Princess cruise ship with 7 dears and 725 cases in a population of 3,700. In both cases, we try to use the distribution of population and of deaths to estimate to the correct population fatality rate for an average member of the population.

Initially, we lacked information on the age distribution of those on the Diamond Princess but fortunately, Dr. Francesco Zonta from Shanghaitech found a paper with the data that was needed for a proper study of the Diamond Princess (https://www.medrxiv.org/content/10.1101/2020.02.20.20025866v2.full.pdf+html).

The Diamond Princess Cruise ship can be seen as an unintended experiment to infect all the passenger of the with COVID-19. As such it allows us to estimate the population death rate without having to worry about what constitutes a case. There were approximately 1,690 people on the Diamond Princess who are over 65 years old. There were 7 deaths so the COVID-19 population death rate for those over 65 is 7/1,690 or 0.41%.

By comparison, in the USA in 2017 to 2018, about 51,000 people over 65 years old died from influenza out of a population of the same age group of 53,000,000. This gives the influenza population death rate is 51,000/53,000,000 or 0.096% (https://www.cdc.gov/flu/about/burden/2017-2018.htm).

This means that if COVID-19 spreads everywhere like influenza has, it will be 0.41/0.096 = 4.3 times more lethal than flu was to people in the USA over 65 years of age in 2017/18.

Of course, it likely is foolish to extrapolate from a single cruise ship to the entire world but we can think of no other way to estimate population death rate. If conditions on the cruise ship were particularly good for transmission of the disease or the older people on the cruise were particularly unhealthy, the difference between COVID and flu would be less.

I expect (or perhaps hope) there will be a vaccine soon and I expect coronavirus to end-up like influenza, infecting almost everyone and about as dangerous. I have heard from an experts that Coronavirus does not have influenza virus' ability to mutate each year so it will likely be as smaller and smaller threat as we all become immune to it less and less severe. It may also behave like the , other corona viruses that are now common colds.

Multiple Superimposed Epidemics

Fig. 13 shows plots of the natural log of the number of cases against date for many different countries selected to have at least 100 cases and five deaths. This was done to put all data on a similar scale and thus simplify the task of examining what is become a 'big data' problem. We were able to identify several interesting situations. Another advantage of the log plots is that exponential growth is to 'tamed' to straight line growth. **Fig. 13A** shows that China non-Hubei has a beautifully smooth curve (red line), the line for China Hubei (olive line) has the same shape but has a small bump between 26-Jan. and 29-Jan. While this initially seemed like noise, it now seems to be an additional outbreak in Hubei. The data from the Diamond Princess is like that of China non-Hubei but is noisier due to smaller numbers. The curve for South Korea is most interesting in that there seem to have been three independent outbreaks starting on 22-Jan., 29-Jan., and 19-Feb.; all were controlled to lead to the flat plateaus that made the new outbreaks easy to see.

Excited by how much information these log plots of the smoothed data contain, we calculated the differences between successive days, namely *ln(N(t))* and *ln(N(t-1))* which high school math teaches is simply *ln(N(t)/N(t-1))*, the natural log of the Change Ratio we have been using since the first analysis. **Fig. 14** shows some of this behavior. Most exciting is that the smooth drop of ln(Change Ratio) seen the very clean data for China non-Hubei seems to be similar to that seen in many other cases but there is a complication of increases do the other outbreaks.

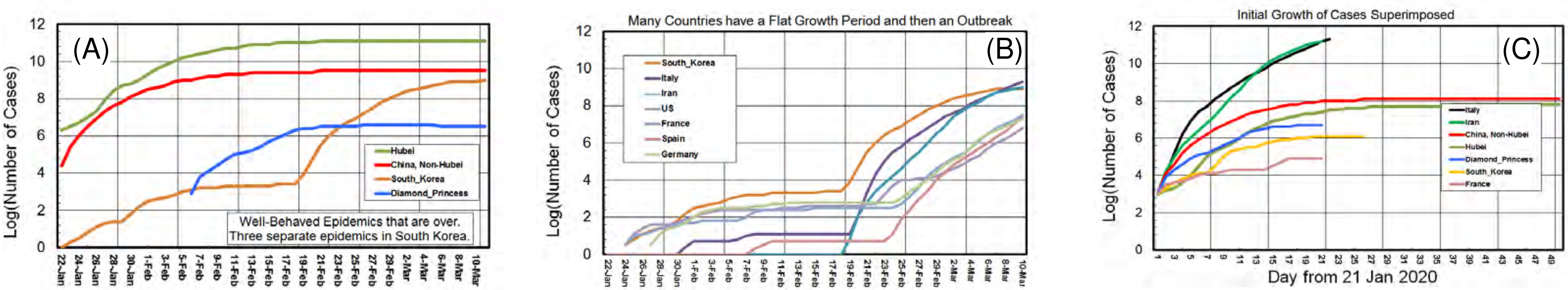


Figure 13. Showing how the natural log of the number of cases of different countries compare. (A) Shows the four outbreaks that are close to ending, Hubei, China Non-Hubei, South Korea and the Diamond Princess. South Korea is unusual in that there seem to have been three independent outbreaks. (B) Shows that many countries have a small number of cases but that for all shown here, there was then a sudden initial jump to show the start of an outbreak: for example for Italy this jump occurred on 19-Feb. whereas in Spain it occurred on 23-Feb. (C) Compares the initial growth in number of cases superimposed to start at same time (Day=1) on the x axis and at he same level value of *log(N)* = 3.0 ($N=e^3=20$) on the y axis. Over a 21 day period these initial events have grown following curves similar to that of China, non-Hubei, our one perfect example. We need to understand what function is being followed but must first finalize and release this report.

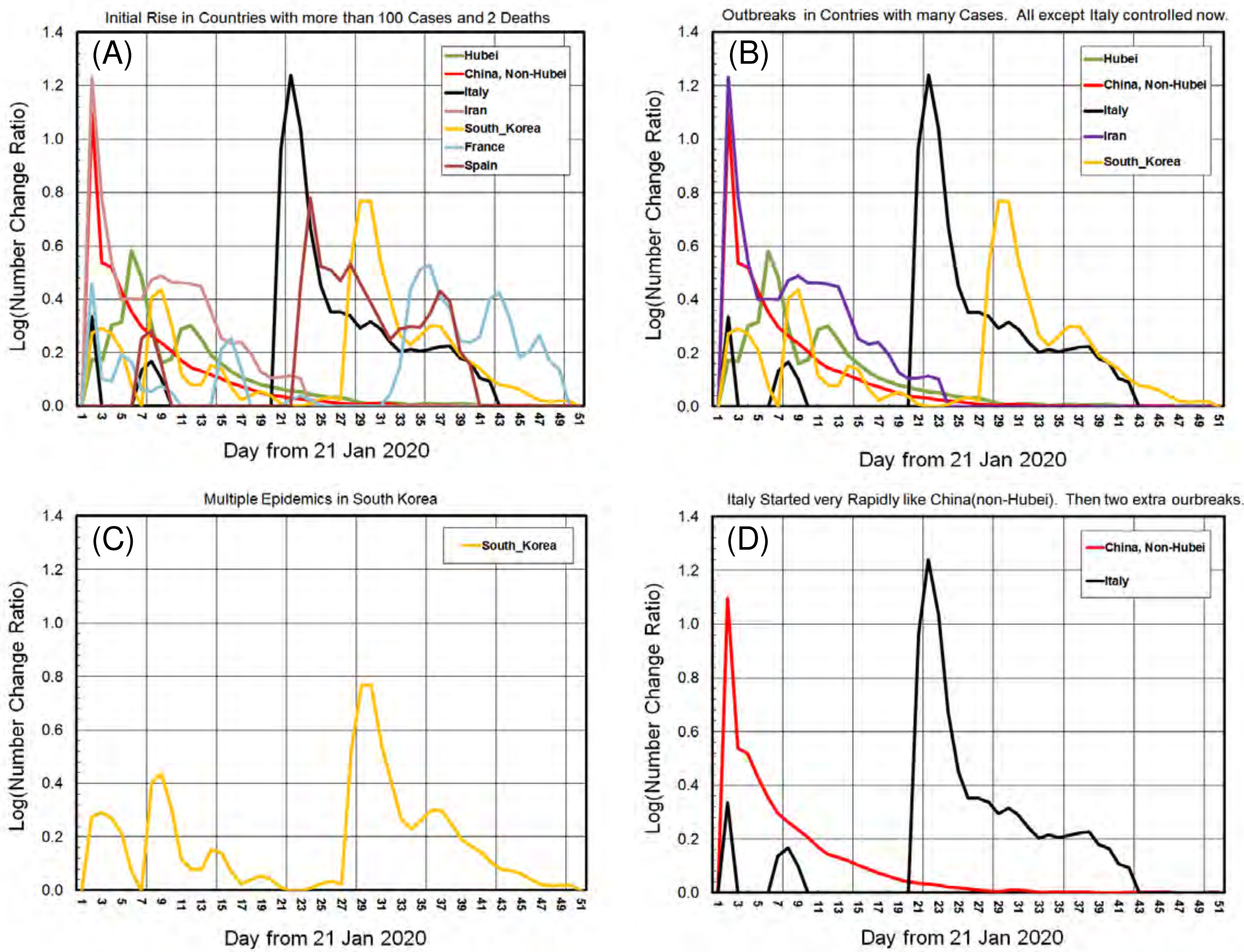


Figure 14. Comparing the natural log of the Change Ratio for cases n different countries. There is a huge amount of detailed and likely useful information that I was sure was due to noise when I first looked at the data from Hubei. Now after LOWESS smoothing it seems much more relevant. (A) Shows the initial spikes of the Change Ratio that marks the start of an outbreak. The separate peaks for the multiple outbreaks in South Korea (orange line), Hubei (olive lone) and in France (pale blue line are clear). (B) shows that the initial high peak relaxes very smoothly for China, non-Hubei and that other locations tend to follow this same relaxation interrupted with added peaks that are sometimes very broad (Iran and Italy). (C) shows the five independents peaks for South Korea corresponding to independent outbreaks. (D) compares the best behaved situation from China, non-Hubei to the worst from Italy.

Different Ways to Measure a Case May Not Matter

The death Rate in Italy is 6.7% which is more than eight times higher than the death rate in South Korea but **Fig 15A** shows that cases in Italy and South Korea grow at exactly the same rate in that the curves can be superimposed on the log cases plot (**Fig. 13**) for a 17 day period for 21-Feb to 8 Mar. This suggests that while Italy is likely missing 7/8ths of the cases South Korea finds, a case as defined by Italy is as a good indicator of the case found in South Korea. This is equally true for cases found in other countries (**Fig. 15C**) and suggests that the lack of testing should not hamper control provided people feeling even slight symptoms self-quarantine.

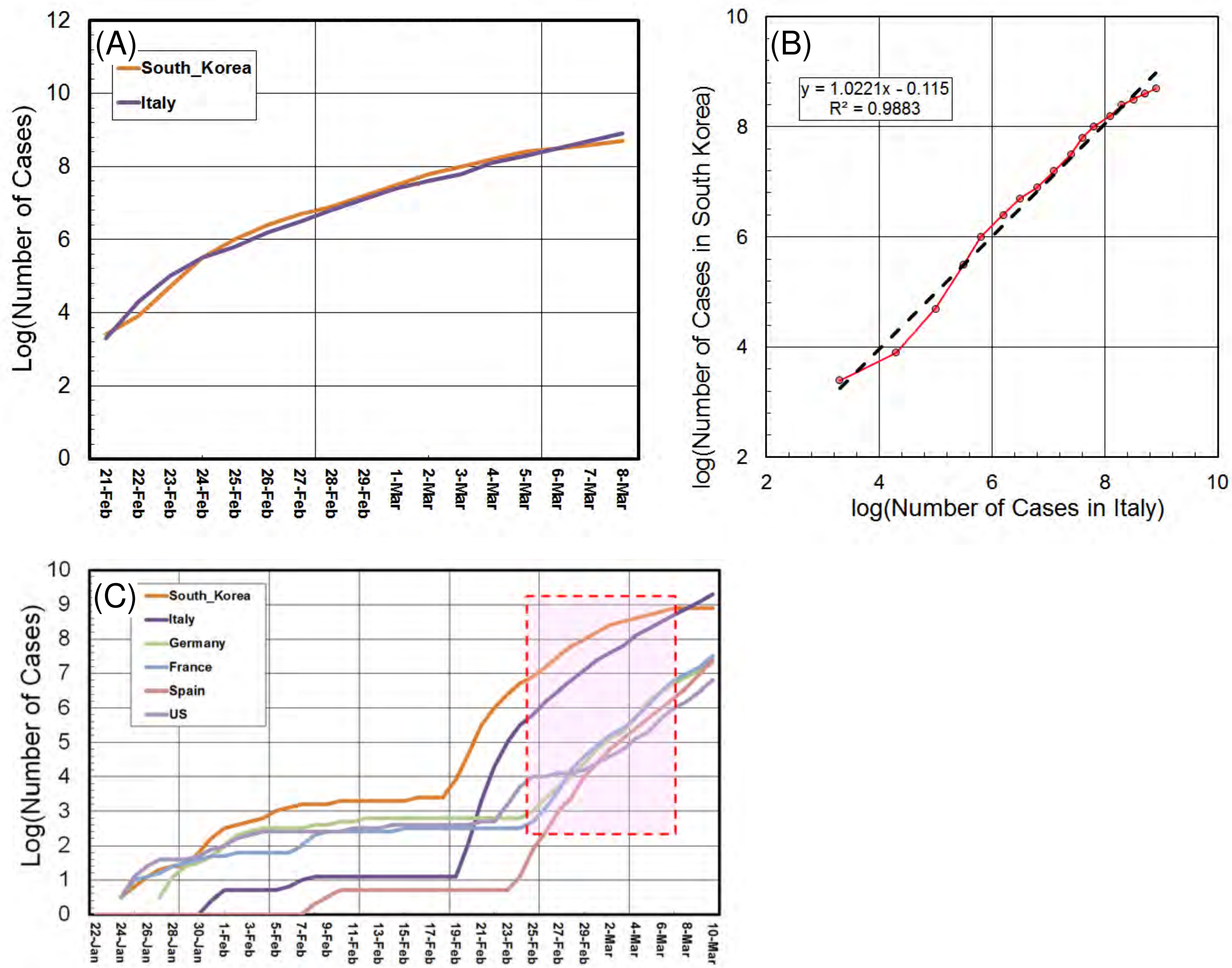


Figure 15. (A) shows how cases in Italy and South Korea have very similar rates of change on a log plot. (B) confirms this by plotting log Cases in South Korea against log Cases in Italy for the 17 day period of overlap. The straight line fit is very good with a R2 of 0.9883 (correlation coefficient of 0.994). The slope is very close to the expected value of 1.0. (C) copied from Fig. 13 highlights in the pink box the period when the differs countries listed gave similar rats of case growth (South Korea, Italy, Germany, Spain and the US).

Wide Range of Case Fatality Ratios (Death Rates)

Why do the death rates of the countries we are focusing on differ so much ranging as the do from 0.14% to 6.72%. Age of population had been considered a reason but as Table 16 shows there seem to be little connection.

Assuming that we are dealing with the same virus with the same intrinsic R0 and time infectious, what can be happening? A very good article about this appeared in today's NYT (appended). There was also a good article in Time Magazine relating to Italy and the role that the low influenza vaccinations may play (also appended).

In my group WhatsApp discussion today with Dr. João Rodrigues and others, the following emerged. "The SE Asian countries are better prepared to handle these sorts of outbreaks. The SARS epidemics and influenza (H1N1, H5N1, etc) led to a generally high public awareness for infectious diseases and transmission as well as the permanent installation of temperature sensors in airports in China. The culture of even slightly sick people always wearing masks in public probably helped, but more importantly either strong governments (China) or very well-organized ones (Singapore, Taiwan, Hong Kong) with their public listening to and supporting them. The US and Europe is more liberal and confidence in government actions is lower, so people act more recklessly." This ignores the low death rates in Germany and Japan, but these countries also share a tradition of belief in their governments.

TO BE CONTINUED…

Country	Death Rate %	% Over 65
Germany	0.14	21.46
Switzerland	0.61	18.62
China	0.82	10.92
South_Korea	0.82	14.42
Belgium	1.01	18.79
Netherlands	1.01	19.20
United_Kingdom	1.66	18.40
France	2.02	20.03
Australia	2.24	15.66
Hong_Kong	2.24	16.88
Japan	2.47	27.58
Spain	2.47	19.38
US	2.47	15.81
Iran	4.50	6.18
Italy	6.72	22.75

Table 16: Showing Case Fatality Ratios or Death Rates for data of 13-Mar. The data on percent of population over 65 are for 2018 and taken from https://data.worldbank.org/indicator/SP.POP.65UP.TO